

Underway Data Processing

Parent Category: References (/references.html)

Category: Methods (/about-calcofi/methods.html)

© Last Updated: 11 March 2016

Underway Data Processing Methods

1. Navigate to Underway Data folder and join into a single file, do so using Advanced File Joiner or similar:
http://download.cnet.com/Advanced-File-Joiner/3000-2094_4-169639.html (http://download.cnet.com/Advanced-File-Joiner/3000-2094_4-169639.html)
 - a. CalCOFI collects its own data while aboard SIO ships (New Horizon). These are saved in the Events folder of the appropriate cruise. A single file name is similar to: 0901NHMETEV_06012009.csv. Although each filename is labeled by date, date data may not correlate. This is of no particular consequence, just be aware of it.
 - b. When a cruise is aboard a NOAA ship, it is normal to receive a CD-ROM at the cruise with a variety of SCS data. These data formats are highly variable by ship and by year. It will usually take a combination of files to find all that you will require (i.e., one file type for TSG, one for fluorescence, one for lat long). Just merge each file type, then cut and paste columns into the correct order. Make all SCS data available on the web.
 - c. The resultant file becomes the RAW UNDERWAY DATA for that cruise. This file contains all unedited data from the Shipboard Computer System (SCS). This file may include navigational, meteorological, oceanographic, and other mechanical data. In your local working directory (e.g., ../UnderwayData/2009/0911NH) save raw data in tab delimited ASCII format and name as YMMSS_UnderwayRaw.txt (e.g., 0901NH_RawUnderway.txt)
 - d. *Links to SCS information
 - i. <http://oceanexplorer.noaa.gov/technology/tools/scs/scs.html>
(<http://oceanexplorer.noaa.gov/technology/tools/scs/scs.html>)
 - ii. http://www.unols.org/meetings/2006/200610inm/SessionIIIa/SessionIIIa_Shields.pdf
(http://www.unols.org/meetings/2006/200610inm/SessionIIIa/SessionIIIa_Shields.pdf)
 - iii. <http://sdsioa.ucsd.edu/> (<http://sdsioa.ucsd.edu/>)
2. Open merged file and crop data columns to include the following, in order and with the following column headers:
 - a. DATE.UTC, TIME.UTC, LA, LO, ST, SA, FL, FI (Date(UTC), Time(UTC), Lat, Lon, Temp, Salinity, Fluor, Pump)
 - i. *Note - There is the possibility that power gets toggled on/off to underway sensors during the cruise. This results in changes in the total number of data columns (i.e., the first days may have 100 columns of data and the next few only 98). Attention is needed to make sure that when cropping data that the correct data is sub sampled.

ii. ****Note** - Different ships have different underway column headers/variable designations/configurations. Consult Underway Data Formats (link (<http://calcofi.org/data/underway/151-underway-data-format/261-underway-data-format.html>)) for more information.

3. Delete rows with:

a. pump speed not within than manufacturer's specifications

i. *Doubling the flow rate from what we recommend will have at least one main problem. This flow rate will probably be too high for the conductivity cell and overtime you can see a degradation of the conductivity readings due to this flow rate. What can happen is that the high flow rate may slowly remove the platinum on the electrode within the conductivity cell, especially in areas with a lots of debris and particulate in the water. This will change the area of the conductivity cell which can cause the readings to be high of correct. I am unsure how long this process would take but it is something we have seen before. I really suggest using the recommended flow rate for this instrument. ~SBE Technical Support*

b. no spatial coordinates

c. erroneous values

i. ***Note** - There is the possibility that one or more columns may be missing some or all of the data. There are several workarounds depending on the missing variable. Consult WTF Moments (link) for more information.

ii. ****Note** - With upwards of 100,000 rows of data it is inevitable that most bad values will go unnoticed at this point. Subsequent data processing steps (i.e., plotting) will make it easier to find errors and to remove them. Just try to find the large and obvious discrepancies.

4. Save cropped file as 0901NH_UnderwayCrop.txt in same folder as the raw file.

5. It is necessary to compare and calculate correction regressions based on the comparison of CTD to Underway Data. Navigate to the final CTD Downcast .csv file folder and join all csvs into a single file (as in Step 1). These csv files are the ones with IEH bottle data added to them. In your local working directory (e.g., ../UnderwayData/2009/0901NH) save raw data in tab delimited ASCII format and name as YMMSS_CTDd.txt (e.g., 0901NH_CTDd.txt). If the YMM_CTDd.csv file is available open in Excel and crop columns accordingly (#7).

6. Open file with MS Excel and select the first column 'A' and Text-to-Column the Data with comma delimiter.

a. ***Note** - DO NOT save this file as YMMSS_CTDd.txt while in Excel. A host of things can go wrong. Data will be Cut and Pasted into a new file shortly. If you must you can save it as some tempfile.xlsx.

7. Crop data columns to include the following, in order and with the following column headers:

a. Date_Time_UTC, Lat_Dec, Lon_Dec, Temp0, Temp1, BTL_Depth, BTL_Salt, BTL_Ch1

i. *Note - If Underway Data was only available with Date and Time in PST select Date_Time_PST instead.

8. Sort Ascending (Smallest to Largest) based on BTL_Depth and Delete all depths greater than 2m, all rows with no BTL_Depth values, and all Header rows except for one which needs to be cut and pasted back at the top after we are finished sorting.
9. Sort Ascending based on Date_Time.
10. Delete any rows that are missing any information.
11. Delete any duplicate stations. You can tell if two subsequent rows are from the same station by the time, delete the deeper.
12. Make sure Header row is at the top.
13. After the Temp 1 column, insert 2 additional blank columns.
 - a. Average the two Temp columns into the first blank column, then copy and paste 'values only' into the second. Delete the Temp0, Temp1, and first column. Header the remaining column as TempAvg.
14. Add three extra headers/columns at the end of the existing
 - a. ..., ST, SA, FL, FI (... , T, S, Fluro, Pump)
15. Open a previously completed copy of YMMSS_CTDvsUnderway.xlsx. Select, Copy, and Paste all rows and columns from the sorted and corrected ...CTDd.txt file into the recently opened xlsx file. Save and RENAME the xlsx file accordingly (i.e., NEW YMMSS). Make sure no old data lingers in the spreadsheet as this would cause errors in the fit of the regression.
 - a. *Note - We reuse the xlsx file because it already has multiple graphs set up. You will of course need to resize the data sets for each graph as necessary, and rename the graph titles. But if you followed the above steps the columns will be in the right places and everything should fall into place.
16. Open the ...UnderwayCrop.txt in MS Excel, Tab delimited. Arrange both files in Tile View.
17. Using this new ...CTDvsUnderway.xlsx file's Date_Time as a guide, scroll down the ...UnderwayCrop.txt and find a corresponding date and time. There will be more than one row of data for the time you are trying to match. Try not to bias the selection of the Underway data and just pick the first one you see. Select the four columns, Temp, Salinity, Fluor, Pump, from UnderwayCrop and Copy them, and Paste them to the xlsx file in the last four columns ST, SA, etc. Do this for all the rows in the CTDvsUnderway file.
 - a. Because the number of rows will change from cruise to cruise, the graph data source will need to be adjusted accordingly. This is important as the linear regression equations shown on the graphs are used to correct the Underway Data in subsequent steps. Save file.

18. For export to MATLAB it is necessary to Cut the xlsx data (numbers no graph info) and paste it into a new UltraEdit .txt file. Name it accordingly (YYMMSS_CTDvsUnderway.txt). This is done to remove any .xlsx formatting that may cause MATLAB to choke on import.
19. Start Matlab and load a previously completed .m file as a template (e.g., 0901NH.m). The .m file will have incorrect variable names and notes pertaining to that specific cruise. Change and update accordingly. Also make sure you are working in your current underway folder.
 - a. Import UnderwayCrop.txt and follow along the notes in the script for error checking and saving data as a .mat file.
 - b. Import ...CTDvsUnderway.txt. There will be an error with this as the Date_Time will not ingest properly. Ignore it, the date is not important for plotting unless GPS has failed and you need the data and time to correlate with the Event Log GPS to get some coordinates to make these maps.
 - c. An important step to be aware of is toward the end of the script it asks you to use dlmwrite function to export the final underway data product.
20. The template m-file will guide you through the entire process. When you get to the regression corrections take the values from the xlsx file graphs that you generated earlier. When you make the Matlab plots they should print into your working folder.
21. Open a previous yymmplot.ai file in Adobe Illustrator and the first plot from Matlab (i.e., t0901NHpath.eps). Save the .ai file with the current cruise YYMM.
 - a. In dealing with the .eps file from Matlab you will find that all we really need is the color path and the gradient bar frame and numbers. Ungroup and delete all you don't need (coastline, grid, tick marks, gradient color, etc.)
 - b. Since the path is thousands of small circles, select them all, group them, then copy them to the clipboard. Delete the preexisting path in the .ai file and make sure you paste the new one in the correct .ai layer and resize to fit the station layout / cruise path. This isn't an exacting process, just make it fit best you can. Make sure to lock/unlock layers as you go to prevent the pasting / editing in the wrong place.
 - c. Go back to the eps and copy the gradient bar box, tickmarks, and numbers. Since each cruise's graph has a different range of values these gradient bars change. The colors stay the same (if you followed the Matlab m-file correctly) but we need to make the new numbers fit the template. Delete the old box and numbers and paste in the new and resize accordingly. Resizing will skew the font so once everything is in the right place check the font setting and set to 18pt Myriad Pro with 100% font stretching in both directions. You will also have to change the decimals and numbers a bit as Matlab tend to separate the numbers after the decimal in eps format. Just delete the end and retype the number using the left half.
 - d. Unlock and Edit the Title layer to the correct cruise designation and dates and title.
 - e. In Illustrator to save as pdf "save a copy", to save as png Use Export (white background, medium resolution, anti aliased) . Open each png in Photoshop and shrink to 75 pixels keep proportionality and rename with a "_t" at the end for the web thumb.
31. Upload all data and images to calcofi.org FTP Underway_Data directory.

32. Edit calcofi.org Joomla site with new underway data accordingly.